

**FACULDADE DE ENGENHARIA DA UNIVERSIDADE DO PORTO**

# **Um CRIS para uma Unidade de Investigação**

**Daniel Moreira Couceiro**



Mestrado Integrado em Engenharia Informática e Computação

Orientador: Gabriel de Sousa Torcato David

24 de Julho de 2017



# **Um CRIS para uma Unidade de Investigação**

**Daniel Moreira Couceiro**

Mestrado Integrado em Engenharia Informática e Computação

Aprovado em provas públicas pelo Júri:

Presidente: Maria Cristina de Carvalho Alves Ribeiro

Arguente: Maria Benedita Campos Neves Malheiro

Vogal: Gabriel de Sousa Torcato David

24 de Julho de 2017



# Resumo

Este trabalho insere-se no contexto das Unidades de Investigação (UI), sendo que o caso de estudo principal é o Instituto de Engenharia de Sistemas e Computadores, Tecnologia e Ciência (INESC TEC). Este tipo de organizações exige uma gestão cuidada da sua informação e dos seus recursos. Tipicamente, as pessoas participam em projetos que podem envolver entidades externas, tais como financiadores, parceiros ou destinatários. Destes projetos são produzidos diferentes tipos de resultados como relatórios, protótipos, *datasets*, publicações ou patentes. Todo este conjunto de dados é utilizado a nível interno para a avaliação individual e das diferentes estruturas orgânicas que compõem a organização e a nível externo para a divulgação de resultados e da atividade. O sistema de informação que contempla estas entidades e as relações entre si é designado por *Current Research Information System* (CRIS) e tem sido alvo de normalização. Uma das normas, proposta pela União Europeia e desenvolvida pela euroCRIS, é conhecida por *Common European Research Information Format* (CERIF). A norma CERIF formaliza as entidades no âmbito da investigação e as suas relações, promovendo desta forma a interoperabilidade entre diferentes sistemas.

Com este projeto pretende-se desenhar um CRIS e contribuir para a sua implementação no INESC TEC, tendo em conta as orientações formalizadas na norma CERIF e os sistemas de informação existentes. Este sistema deverá incluir a gestão de projetos de investigação, suas equipas, etapas e resultados. Para além disso, dever-se-á relacionar os projetos com os investigadores envolvidos, repositórios de publicações e dados de investigação e a produção de indicadores segundo perspetivas interna e externa.



# Abstract

This project is inserted in the Research Units (UI) context, whereas the main case study is *Instituto de Engenharia de Sistemas e Computadores, Tecnologia e Ciência* (INESC TEC). This type of organizations requires a careful management of its information and resources. Typically, the individuals work in projects which may involve external entities, such as financiers, partners or addressees. These projects produce several types of results which include reports, prototypes, data sets, publications or patents. All of this information is used internally, for the evaluation of the individuals and the various organic structures which compose the organization, and externally, for the divulgation of results and its activity. The information system that contemplates these entities and the relations among them is commonly known as Current Research Information System (CRIS) and has been target of normalization. One of the standards, proposed by the European Union and developed and maintained by euroCRIS, is named Common European Research Information Format (CERIF). This standard formalizes the research related entities and its relations, thus emphasizing the interoperability among different systems.

With this project it is expected to design a CRIS and contribute to its implementation in INESC TEC, while having in mind the formalized orientations in CERIF and the existing information systems. This system should include the management of research projects, its teams, milestones and results. Furthermore, it should relate the projects with the involved researchers, publications and research data repositories and comprise the production of metrics along internal and external perspectives.





# Agradecimentos

Começo por agradecer ao meu orientador, Gabriel David, pela oportunidade de trabalho neste projeto e pelo apoio prestado durante a realização do mesmo.

Agradeço também aos meus colegas do INESC TEC pelo ambiente de trabalho proporcionado e pela ajuda em diversas situações: Fábio Alves, Isabel Macedo, José Carlos Sousa, Mafalda Landeiro e Rúben Moreira.

Finalmente, agradeço aos meus familiares que, de uma forma ou de outra, me motivaram e tornaram esta etapa mais fácil.

Daniel Moreira Couceiro



*“Fahre fort, übe nicht allein die Kunst, sondern dringe auch in ihr Inneres; sie verdient es.  
Denn nur die Kunst und die Wissenschaft erhöhen den Menschen bis zur Gottheit.”*

Ludwig van Beethoven



# Conteúdo

<b>1</b>	<b>Introdução</b>	<b>1</b>
1.1	Contexto/Enquadramento . . . . .	1
1.2	Motivação e Objetivos . . . . .	2
1.3	Estrutura da Dissertação . . . . .	2
<b>2</b>	<b>Norma CERIF</b>	<b>3</b>
2.1	Apresentação . . . . .	3
2.2	Modelo de dados . . . . .	3
2.3	CERIF XML . . . . .	5
2.4	Resumo . . . . .	5
<b>3</b>	<b>Revisão Tecnológica</b>	<b>7</b>
3.1	Soluções comerciais . . . . .	7
3.1.1	Converis . . . . .	7
3.1.2	Pure . . . . .	7
3.2	DSpace-CRIS . . . . .	8
3.2.1	Apresentação . . . . .	8
3.2.2	Funcionalidades . . . . .	8
3.3	PTCRIS . . . . .	8
3.3.1	Apresentação . . . . .	8
3.3.2	Contexto . . . . .	9
3.3.3	Ecossistema . . . . .	9
3.3.4	Benefícios . . . . .	10
3.4	Microsoft SQL Server . . . . .	11
3.5	Spring Framework . . . . .	12
3.6	Angular . . . . .	13
3.7	Conclusão . . . . .	14
<b>4</b>	<b>Especificação</b>	<b>15</b>
4.1	Arquitetura geral . . . . .	15
4.2	Projetos . . . . .	16
4.2.1	Especificação . . . . .	16
4.2.2	Modelo de dados . . . . .	17
4.3	Publicações . . . . .	19
4.3.1	Especificação . . . . .	19
4.3.2	Modelo de dados . . . . .	19
4.4	Indicadores . . . . .	20
4.4.1	Modelo de dados . . . . .	21

## CONTEÚDO

4.5	Resumo . . . . .	22
<b>5</b>	<b>Implementação</b>	<b>23</b>
5.1	Servidor Spring . . . . .	23
5.2	Projetos . . . . .	24
5.2.1	Página de pesquisa . . . . .	24
5.2.2	Página de visualização . . . . .	25
5.2.3	Página de edição . . . . .	27
5.3	Publicações . . . . .	28
5.4	Exportação em CERIF XML . . . . .	29
5.5	Resumo . . . . .	29
<b>6</b>	<b>Conclusões e Trabalho Futuro</b>	<b>31</b>
6.1	Conclusão e Satisfação dos Objetivos . . . . .	31
6.2	Trabalho Futuro . . . . .	31
	<b>Referências</b>	<b>33</b>

# Lista de Figuras

2.1	CERIF - Modelo de dados . . . . .	4
3.1	Ecosistema de ciência e tecnologia, disponível em <a href="https://ptcris.pt/sobre-o-ptcris/apresentacao/">https://ptcris.pt/sobre-o-ptcris/apresentacao/</a> . . . . .	9
3.2	Universo CRIS, disponível em <a href="https://ptcris.pt/ecossistema/">https://ptcris.pt/ecossistema/</a> . . . . .	10
3.3	Ranking DB-Engines em 2017-06-20. Disponível em <a href="https://db-engines.com/en/ranking">https://db-engines.com/en/ranking</a> . . . . .	12
3.4	Ranking HotFrameworks em 2017-06-21. Disponível em <a href="https://hotframeworks.com/">https://hotframeworks.com/</a> . . . . .	12
3.5	Arquitetura do Angular. Disponível em <a href="https://angular.io/guide/architecture">https://angular.io/guide/architecture</a> . . . . .	13
4.1	Pilha tecnológica . . . . .	16
4.2	Modelo de dados de projetos . . . . .	18
4.3	Modelo de dados de publicações . . . . .	20
4.4	Modelo de dados de indicadores . . . . .	21
5.1	Página de pesquisa de projetos . . . . .	24
5.2	Identificação do projeto . . . . .	25
5.3	Informação geral do projeto . . . . .	26
5.4	Investigadores no projeto . . . . .	26
5.5	Edição de centros envolvidos no projeto . . . . .	27
5.6	Edição de <i>Work Packages</i> do projeto . . . . .	27
5.7	Edição de bolsiros do projeto . . . . .	28
5.8	Módulo de publicações . . . . .	28

## LISTA DE FIGURAS



# Lista de Tabelas

3.1	Tabela de Comparação entre os CRIS . . . . .	14
-----	--	----

## LISTA DE TABELAS

# Abreviaturas e Símbolos

ACID	<i>Atomicity, Consistency, Isolation, Durability</i>
CAS	<i>Central Authentication Service</i>
CERIF	<i>Common European Research Information Format</i>
CRIS	<i>Current Research Information System</i>
CRUD	<i>Create, Read, Update, Delete</i>
CORE	<i>Computing Research and Education Association of Australasia</i>
C&T	<i>Ciência e Tecnologia</i>
CV	<i>Curriculum Vitae</i>
DBLP	<i>Digital Bibliography &amp; Library Project</i>
FCT	<i>Fundação para a Ciência e a Tecnologia</i>
HTML	<i>HyperText Markup Language</i>
ISI	<i>Institute for Scientific Information</i>
ISSN	<i>International Standard Serial Number</i>
LDAP	<i>Lightweight Directory Access Protocol</i>
OLAP	<i>Online Analytical Processing</i>
ORCID	<i>Open Researcher and Contributor ID</i>
RDBMS	<i>Relational DataBase Management System</i>
RI	<i>Repositório Institucional</i>
TLR	<i>Technology Readiness Level</i>
UI	<i>Unidade de Investigação</i>
URL	<i>Uniform Resource Locator</i>
WoS	<i>Web of Science</i>
XML	<i>eXtensible Markup Language</i>



# Capítulo 1

## Introdução

Este capítulo tem como objetivo apresentar o contexto no qual a dissertação se insere, bem como uma descrição geral do problema. Para além disso, expõem-se a motivação para o desenvolvimento do projeto e os objetivos pretendidos. Finalmente, apresenta-se a estrutura global do documento incluindo um breve resumo de cada capítulo.

### 1.1 Contexto/Enquadramento

As unidades de investigação (UI) representam um pilar fundamental na consolidação de um sistema científico moderno e competitivo. Devem reunir massa crítica adequada à sua missão e promover ambientes criativos onde possam surgir novas ideias e onde os investigadores encontrem as condições adequadas à realização dos seus projetos científicos e ao desenvolvimento da sua carreira. Sempre que aplicável, devem reunir recursos interdisciplinares e multidisciplinares que potenciem a abordagem de problemas complexos e novos desafios sociais<sup>1</sup>. São organizações muito exigentes quanto à gestão da informação que suporta a sua atividade. O conceito principal de trabalho é o de projeto de investigação que envolve entidades externas (tais como financiadores ou parceiros), equipas, etapas e resultados. Os resultados podem ser de diferentes naturezas, como publicações, dados de investigação ou patentes. Para além disso, toda esta informação é utilizada a nível interno, para a avaliação individual e das diferentes estruturas hierárquicas da organização, e a nível externo, para reporte de resultados e da atividade.

O contexto no qual a dissertação se realiza é o do Instituto de Engenharia de Sistemas e Computadores, Tecnologia e Ciência - INESC TEC, um Laboratório Associado que conta com mais de 30 anos de experiência em I&D e Transferência de Tecnologia. Com 6 polos no Porto (sede), Braga e Vila Real, o INESC TEC agrega 13 Centros de I&D e uma Unidade Associada com competências complementares e vocacionadas para o mercado internacional. Agrupa mais de 650

---

<sup>1</sup>FCT — Unidades de I&D, disponível em <https://www.fct.pt/apoios/unidades/index.phtml.pt>

investigadores, entre os quais mais de 270 doutorados, constituindo um núcleo denso de desenvolvimento científico e tecnológico em Portugal, com uma presença internacional assinalável. A sua atividade visa a Investigação Científica e o Desenvolvimento Tecnológico, mas também a Consultoria e Formação Avançada, bem como a Transferência de Tecnologia e o Lançamento de Novas Empresas de Base Tecnológica<sup>2</sup>.

### 1.2 Motivação e Objetivos

Numa unidade de investigação, o tipo de informação que é necessário gerir é muito diversificado. Assim, é comum optar-se por ferramentas que resolvam um problema específico de urgência mais imediata de cada vez. Isto resulta, tipicamente, em informação replicada, não normalizada e sem ligação entre os vários subsistemas, tornando-os difíceis de gerir e manter. Para além disso, pode também não ser possível registar, de forma expedita, alguns dados desejáveis, sobretudo quando relacionam vários subsistemas diferentes. Para contornar este inconveniente, existem várias soluções comerciais e *open source* disponíveis. Este tipo de soluções, sistemas de informação que gerem a atividade relacionada com a investigação, são denominadas *Current Research Information System* (CRIS). A nível internacional (mas sobretudo europeu) o principal *standard* que normaliza as entidades dos CRIS e as relações entre si é a norma CERIF.

O principal objetivo desta dissertação é projetar um CRIS relativo a uma Unidade de Investigação inserida num contexto universitário. Para além disso, um outro objetivo é contribuir para uma implementação no INESC TEC. Para isso, deve ter-se em consideração a sua atual arquitetura de sistemas de informação e as recomendações formalizadas na norma CERIF.

### 1.3 Estrutura da Dissertação

Para além da introdução, este documento contém mais 5 capítulos. No capítulo 2 é descrita a norma CERIF mais detalhadamente. No capítulo 3 apresentam-se tecnologias existentes no contexto dos CRIS e seleção de *software* relevante para a execução do projeto. No capítulo 4 expõem-se os requisitos do sistema, incluindo uma visão geral da arquitetura e os principais modelos de dados. No capítulo 5 são referidos os principais detalhes de implementação, bem como uma descrição da plataforma desenvolvida. No capítulo 6 indicam-se as principais conclusões a reter da realização desta dissertação. Apresenta-se também o grau de satisfação dos objetivos inicialmente propostos, assim como possíveis pontos de trabalho futuro.

---

<sup>2</sup>Apresentação do INESC TEC, disponível em <http://www.inesctec.pt/apresentacao>

## Capítulo 2

# Norma CERIF

Este capítulo tem como objetivos introduzir a norma CERIF e descrever os principais conceitos e funcionalidades subjacentes.

### 2.1 Apresentação

A euroCRIS é uma associação sem fins lucrativos que junta peritos em informação de investigação em geral e, particularmente, sistemas de informação de investigação (CRIS). A organização tem mais de 200 membros e uma das principais atividades é o desenvolvimento do modelo de dados internacional para informação de investigação denominado CERIF: *Common European Research Information Format*<sup>1</sup>.

A norma CERIF tem como principal objetivo estabelecer um padrão internacional para a organização dos sistemas CRIS e dos dados aí contemplados, com vista a promover a interoperabilidade entre diversos sistemas de informação e conseguir responder a questões, como por exemplo:

- Por que entidades é o projeto X financiado?
- Em quantos projetos FP7 participa a organização Y?
- Quantas publicações resultaram do projeto Z?

### 2.2 Modelo de dados

A norma CERIF define vários tipos de entidades, concetualizados em tabelas relacionais típicas: com colunas para os atributos e linhas para as instâncias. A estrutura dos dados é normalizada de modo a evitar replicação de dados e, consequentemente, problemas de inconsistência. Deste modo, a informação relativa a um conceito pode estar distribuída por várias entidades [Jef10].

---

<sup>1</sup>What is euroCRIS? | euroCRIS, disponível em <http://www.eurocris.org/what-eurocris>

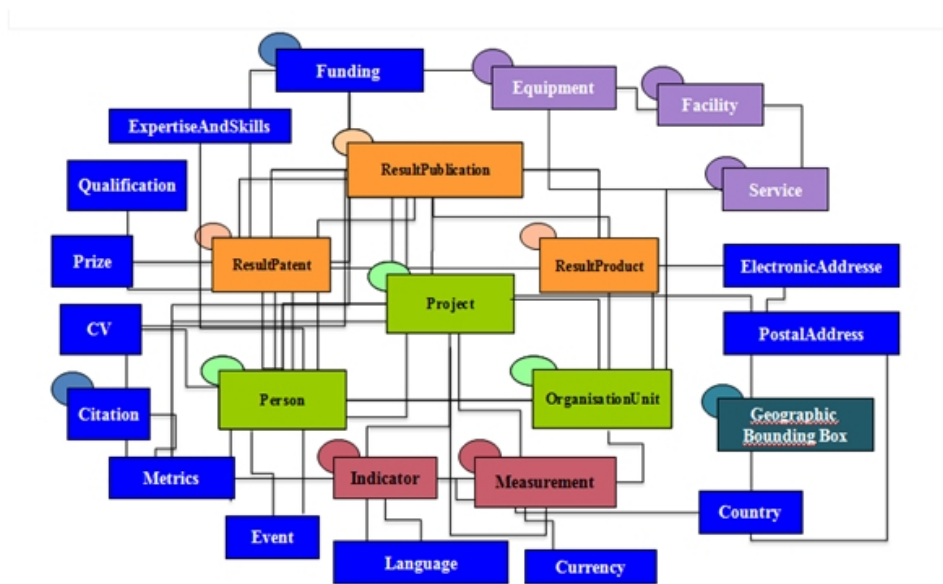


Figura 2.1: CERIF - Modelo de dados

O modelo CERIF é estruturado conceitualmente em tipos e funcionalidades. Os tipos incluem: entidades base, de resultados, secundárias e de ligação. As funcionalidades correspondem a campos multilingue e semântica [Jö10]. Mais especificamente:

- **Entidades base:** As entidades base incluem *Person*, *OrganizationUnit*, *Project* e correspondem aos retângulos de fundo verde na figura 2.1;
- **Entidades de resultados:** As entidades de resultados são representadas por retângulos de fundo laranja na figura 2.1, englobando *ResultProduct*, *ResultPatent* e *ResultPublication*;
- **Entidades secundárias:** Permitem a representação do ambiente de investigação e correspondem, na figura 2.1, aos retângulos de fundo azul;
- **Entidades de ligação:** As entidades de ligação representam as relações entre as entidades base, de resultados e secundárias, sendo um dos pontos fortes do CERIF. Cada uma destas relações inclui um intervalo de validade e uma classificação. Na Figura 2.1 são representadas pelas linhas que ligam as entidades;
- **Funcionalidades multilingue:** O CERIF suporta definição de valores em diferentes línguas para um determinado atributo de uma instância;
- **Funcionalidades de semântica:** A camada semântica do CERIF permite atribuir papéis e intervalos de validade a relações entre entidades. Suporta também a definição de conjuntos de valores fechados para um dado atributo, esquemas de classificação, incluindo até mapeamento de valores entre esquemas.



## 2.3 CERIF XML

O principal meio de interoperabilidade entre sistemas CRIS previsto pelo CERIF é o formato CERIF *eXtensible Markup Language* (XML). Este formato baseia-se na estrutura do modelo conceptual do CERIF. O nome das anotações XML é replicado do nome das entidades relacionais do modelo de dados. Na raiz do documento XML existe um nó CERIF, cujos descendentes diretos são entidades simples (como *cfPers*, *cfProj* ou *cfResPubl*). Contudo, enquanto no modelo relacional CERIF a ligação entre os elementos é realizada através de identificadores internos, no formato CERIF XML os componentes são aninhados hierarquicamente de acordo com a sua ligação [HJD<sup>+</sup>14]. No Código 2.1 é apresentado um exemplo de um documento em CERIF XML.

```

1 <CERIF xmlns="urn:xmlns:org:eurocris:cerif-1.6-2"
2   xsi:schemaLocation="urn:xmlns:org:eurocris:cerif-1.6-2 http://www.eurocris.org/
   Uploads/Web%20pages/CERIF-1.6/CERIF_1.6_2.xsd"
3   date="2015-01-09"
4   sourceDatabase="http://api.examplecris.org">
5   <cfProj>
6     <cfProjId>43</cfProjId>
7     <cfStartDate>2009-12-01</cfStartDate>
8     <cfEndDate>2012-11-30</cfEndDate>
9     <cfAcro>OPENAIRE</cfAcro>
10    <cfURI>http://www.openaire.eu/en/home</cfURI>
11    <cfTitle cfLangCode="en" cfTrans="h">Open Access Infrastructure for Research in
      Europe</cfTitle>
12    <cfProj_Fund>
13      <cfFundId>19</cfFundId>
14      <cfClassId>4e664be9-100d-481c-b60f-b62a94078bac</cfClassId>
15      <cfClassSchemeId>a620795c-7015-482e-bdba-43a761b337a1</cfClassSchemeId>
16      <cfStartDate>2009-12-01T00:00:00</cfStartDate>
17      <cfEndDate>2012-11-30T00:00:00</cfEndDate>
18      <cfFraction>0.83</cfFraction>
19      <cfAmount cfCurrCode="EU">5027379.0</cfAmount>
20    </cfProj_Fund>
21    <cfProj_Pers>
22      <cfPersId>195</cfPersId>
23      <cfClassId>9100acf1-f01a-4b11-8675-2e7e97624b4e</cfClassId>
24      <cfClassSchemeId>24460a7c-1af9-4941-b603-e6f5a6997be9</cfClassSchemeId>
25      <cfStartDate>2009-12-01T00:00:00</cfStartDate>
26      <cfEndDate>2012-11-30T00:00:00</cfEndDate>
27    </cfProj_Pers>
28  </cfProj>
29 </CERIF>

```

Código 2.1: Exemplo de um documento CERIF XML

## 2.4 Resumo

Neste capítulo começou-se por fazer uma breve apresentação da norma CERIF e dos seus principais objetivos. Em seguida, introduziu-se o seu modelo de dados e a forma como as entidades

## Norma CERIF

estão organizadas e relacionadas. Finalmente, foi abordado o principal mecanismo de interoperabilidade entre diferentes sistemas CRIS previsto por esta norma: o CERIF XML.

## Capítulo 3

# Revisão Tecnológica

### 3.1 Soluções comerciais

#### 3.1.1 Converis

O Converis, da Thomson Reuters, é um sistema de gestão da informação de investigação que permite a gestão completa do ciclo de vida da investigação. Parte das funcionalidades do Converis consiste no suporte da norma CERIF, bem como nas facilidades de registo dos resultados da investigação, que é suportado pelo acesso às bases de dados PubMed e Web of Science [vLvWW16].

Para além disso, permite a interoperabilidade com inúmeros sistemas, tais como

- Servidores de *login* (Shibboleth, Kerberos, *Lightweight Directory Access Protocol* (LDAP), *Central Authentication Service* (CAS))
- Sistemas de Recursos Humanos (PeopleSoft, Resource Link)
- Sistemas de Contabilidade e Finanças (PeopleSoft, Agresso, Oracle, Raindance)
- Repositórios Institucionais (DSpace, EPrints, Fedora, Digital Commons)
- Bases de dados de publicações externas (*Web of Science* (WoS), Scopus, PubMed, *Open Researcher and Contributor ID* (ORCID))

#### 3.1.2 Pure

O Pure, pertencente atualmente à Elsevier, é um sistema CRIS que fornece aos investigadores a possibilidade de armazenar, administrar e publicar dados de investigação através de um único sistema. Este sistema agrega informação de fontes interna e externa. É possível inserir manualmente publicações ou obtê-las automaticamente de fontes como Scopus, Web of Science ou arXiv.

Algumas dos principais módulos fornecidos incluem a geração automática de CV, de acordo com a finalidade desejada, e importação de dados de publicações e citações a partir de fontes externas, como Scopus e Web of Science<sup>1</sup>.

<sup>1</sup>Features - Pure | Elsevier, disponível em <https://www.elsevier.com/solutions/pure/features>

## 3.2 DSpace-CRIS

### 3.2.1 Apresentação

De uma maneira geral, existem 2 tipos de plataformas que suportam a gestão de informação de investigação: os já mencionados CRIS e os Repositórios Institucionais (RI). Embora grande parte das funcionalidades e objetivos sejam comuns, existem várias diferenças entre ambos: enquanto os CRIS são mais focados na atividade de investigação em geral, os IR enfatizam a coleção de resultados de investigação, comumente publicações e dados de investigação (*datasets*).

O DSpace é uma aplicação *open source* largamente utilizada para construir repositórios digitais, sendo, desta forma, uma tecnologia que se encaixa na vertente de RI. Adicionalmente, foi desenvolvida uma extensão para esta ferramenta designada DSpace-CRIS que, para além das funcionalidades de repositório, inclui ainda funcionalidades úteis para utilizadores de sistemas CRIS.

### 3.2.2 Funcionalidades

Esta tecnologia é compatível com o CERIF, incluindo uma funcionalidade de exportação em CERIF XML. Alguns componentes do modelo de dados CERIF são já suportados nativamente, tais como as relações temporais e a caracterização semântica. Apesar de o DSpace-CRIS incluir poucas entidades predefinidas, o seu modelo de dados dinâmico e flexível permite facilmente a criação e configuração de novas entidades previstas no CERIF [PBMM14].

## 3.3 PTCRIS

### 3.3.1 Apresentação

O PTCRIS é um programa estruturante para a gestão de ciência e tecnologia que visa promover a integração de vários sistemas de informação de suporte à atividade científica utilizados pelos investigadores, pelos gestores de ciência ou pelo público em geral.

O PTCRIS irá criar um ecossistema aberto, do qual todos os agentes podem facilmente beneficiar e contribuir, com vantagens para todos. Este ecossistema promoverá a excelência da ciência em Portugal, dando-lhe uma maior visibilidade para todos os intervenientes nacionais e internacionais, facilitando os processos de gestão e produção de ciência e inovação, através de acesso facilitado a informação autoritativa, completa e fidedigna. A Figura 3.1 ilustra um ecossistema de ciência e tecnologia<sup>2</sup>.

Um dos objetivos do PTCRIS é facilitar os processos de gestão de resultados de investigação, através da adoção de uma política “introduzir uma vez, reutilizar várias”.

---

<sup>2</sup>Apresentação | PTCris, disponível em <https://ptcris.pt/sobre-o-ptcris/apresentacao/>



Figura 3.1: Ecosistema de ciência e tecnologia, disponível em <https://ptcris.pt/sobre-o-ptcris/apresentacao/>

### 3.3.2 Contexto

O PTCRIS é o resultado de uma recomendação formulada no âmbito de um trabalho de consultadoria comissionado pelo Conselho Diretivo da FCT à equipa técnica da Unidade FCCN (na sequência da fusão desta com a FCT). Este trabalho concluiu que as necessidades de gestão de ciência e tecnologia vão além da missão dos sistemas de informação de suporte ao negócio da FCT (a gestão do financiamento), sendo fundamental uma visão mais abrangente que envolva todos os agentes da comunidade científica nacional e internacional sob a forma de um ecossistema nacional integrado de ciência e tecnologia<sup>3</sup>. A Figura 3.2 apresenta uma visão geral do universo CRIS.

### 3.3.3 Ecossistema

- **Sistema de curricula:** base de dados de investigadores reunindo informação como o seu identificador, formação, cargos, competências, área de investigação, publicações, afiliações profissionais, etc;
- **Base de dados de organizações:** lista das instituições onde se faz investigação e das suas unidades (departamentos, centros, laboratórios, etc), cada uma com respectivo identificador universal;
- **Infraestruturas científicas:** base de dados das instalações e equipamentos científicos e respetiva localização;
- **Indicadores:** conjunto de estatísticas descritivas da atividade de investigação C&T;

<sup>3</sup>Contexto | PTCris, disponível em <https://ptcris.pt/sobre-o-ptcris/contexto/>

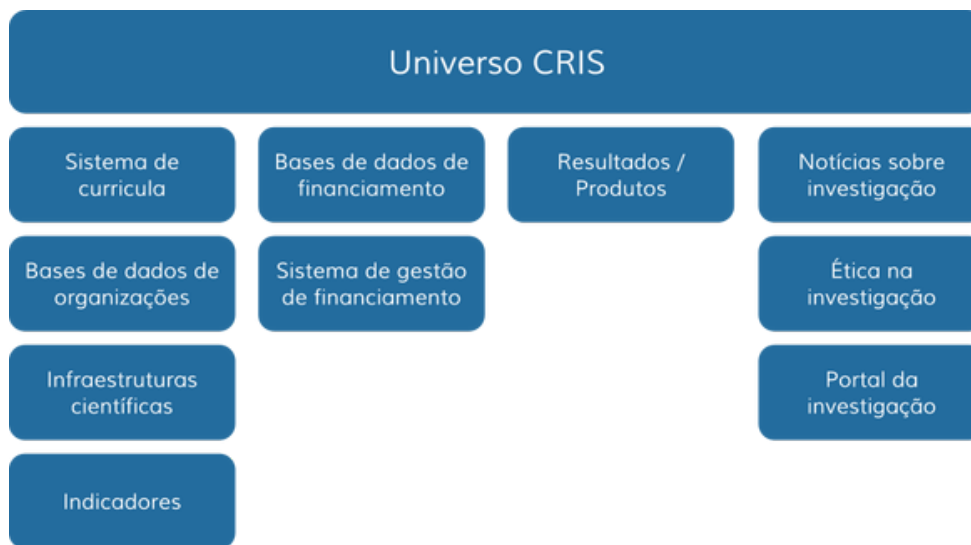


Figura 3.2: Universo CRIS, disponível em <https://ptcris.pt/ecossistema/>

- **Base de dados de financiamento:** listagem dos programas, mecanismos e ações de financiamento das atividades de investigação C&T;
- **Sistema de gestão de financiamento:** plataforma de submissão de candidaturas a financiamento, bem como de gestão dos financiamentos atribuídos e de comunicação com a entidade financiadora;
- **Resultados/produtos:** base de dados de produtos da atividade de investigação C&T sob a forma de dados, publicações (teses, artigos científicos, livros/capítulos de livros, etc), patentes e produtos;
- **Notícias sobre investigação:** agência de notícias relacionadas com a investigação C&T;
- **Ética na Investigação:** informação sobre normas de conduta dos investigadores orientadas por princípios éticos;
- **Portal da Investigação:** base de dados que disponibiliza toda a informação sobre a investigação C&T, ligando investigadores a projetos e resultados/produtos<sup>4</sup>.

### 3.3.4 Benefícios

O PTCRIS irá criar um ecossistema aberto, do qual todos os agentes podem facilmente beneficiar e contribuir, com vantagens para todos. Os benefícios a atingir com a implementação do programa PTCRIS são os seguintes<sup>5</sup>:

<sup>4</sup>Ecossistema | PTCris, disponível em <https://ptcris.pt/ecossistema/>

<sup>5</sup>Benefícios | PTCris, disponível em <https://ptcris.pt/beneficios/>

**Para conselhos de investigação**

- Otimização do processo de financiamento

**Para investigadores**

- Facilidade em aceder a informação relevante
- Facilidade em gerir e reportar

**Para administradores e gestores de ciência**

- Facilidade em medir e analisar a actividade de investigação
- Facilidade no acesso a informações comparativas
- Avaliação

**Para empresas e organizações de transferência de tecnologia**

- Facilidade em descobrir tecnologias e ideias inovadoras
- Facilidade na identificação da concorrência

**Para os media e o público**

- Facilidade no acesso a informação fidedigna, completa e atualizada sobre as instituições onde se pratica investigação C&T, os seus investigadores, respetivos projetos de investigação e os seus resultados

### **3.4 Microsoft SQL Server**

O Microsoft SQL Server é um sistema gestor de bases de dados relacionais (RDBMS) desenvolvido pela Microsoft. Algumas funcionalidades incluem transações com propriedades ACID, integridade referencial e vistas materializadas. De acordo com o Ranking DB-Engines, foi, em maio de 2017, o terceiro sistema deste tipo mais popular (com base no método de cálculo utilizado) conforme é visível na Figura 3.3.

Rank			DBMS	Database Model	Score		
Jun 2017	May 2017	Jun 2016			Jun 2017	May 2017	Jun 2016
1.	1.	1.	Oracle 🏆 🛒	Relational DBMS	1351.76	-2.55	-97.49
2.	2.	2.	MySQL 🏆 🛒	Relational DBMS	1345.31	+5.28	-24.83
3.	3.	3.	Microsoft SQL Server 🏆 🛒	Relational DBMS	1198.97	-14.84	+33.16
4.	4.	📈 5.	PostgreSQL 🏆 🛒	Relational DBMS	368.54	+2.63	+61.94
5.	5.	📉 4.	MongoDB 🏆 🛒	Document store	335.00	+3.42	+20.38
6.	6.	6.	DB2 🏆	Relational DBMS	187.50	-1.34	-1.07
7.	7.	📈 8.	Microsoft Access	Relational DBMS	126.55	-3.33	+0.32
8.	8.	📉 7.	Cassandra 🏆	Wide column store	124.12	+1.01	-7.00
9.	9.	📈 10.	Redis 🏆	Key-value store	118.89	+1.44	+14.39
10.	10.	📉 9.	SQLite	Relational DBMS	116.71	+0.64	+9.92

Figura 3.3: Ranking DB-Engines em 2017-06-20. Disponível em <https://db-engines.com/en/ranking>

### 3.5 Spring Framework

O Spring é uma *framework open source* para programação em Java desenvolvida pela Pivotal Software e que inclui vários módulos<sup>6</sup>. Um desses módulos é o Spring Data cuja função é fornecer um modelo de programação consistente para acesso a dados e que permite contornar algumas dificuldades comuns ao lidar com bases de dados em aplicações, como a gestão de recursos (aquisição e liberação de recursos da base de dados automaticamente) ou a utilização de transações<sup>7</sup>. Na Figura 3.4 apresenta-se o *ranking* HotFrameworks com as tecnologias mais populares para desenvolvimento de *software*.

Framework	Github Score	Stack Overflow Score	Overall Score
ASP.NET		100	100
AngularJS	97	97	97
Ruby on Rails	93	98	95
ASP.NET MVC		94	94
React	100	84	92
Django	90	93	91
Laravel	92	87	89
Angular	89	86	87
Spring	84	91	87
Express	92	82	87

Figura 3.4: Ranking HotFrameworks em 2017-06-21. Disponível em <https://hotframeworks.com/>

<sup>6</sup>Home | Pivotal, disponível em <https://pivotal.io/>

<sup>7</sup>Spring Data, disponível em <http://projects.spring.io/spring-data/>



### 3.6 Angular

O Angular é uma plataforma *open source* desenvolvida pela Google destinada à construção de aplicações *Web front-end* em HTML e ou JavaScript ou TypeScript - uma linguagem de programação desenvolvida e mantida pela Microsoft que é um superconjunto do JavaScript. A *framework* consiste em várias bibliotecas, algumas essenciais e outras opcionais. As aplicações são formadas compondo *templates* HTML com uma sintaxe própria do Angular, que são geridos pelas classes *Component*<sup>8</sup>. Na Figura 3.5 é apresentada uma visão geral da sua arquitetura.

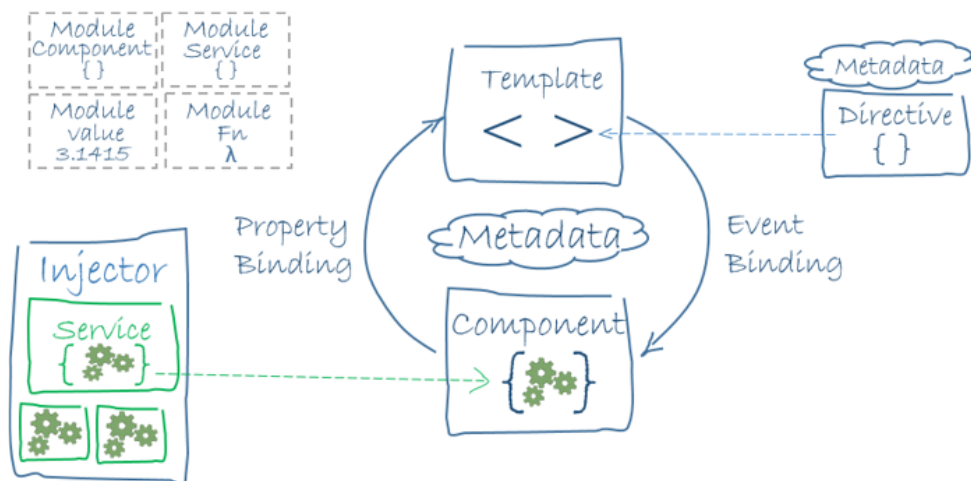


Figura 3.5: Arquitetura do Angular. Disponível em <https://angular.io/guide/architecture>

<sup>8</sup>Angular - Architecture Overview, disponível em <https://angular.io/guide/architecture>

### 3.7 Conclusão

Este capítulo começou por apresentar as soluções no contexto dos CRIS. Inicialmente, foram mencionadas as soluções comerciais mais proeminentes - o Pure e o Converis - e, em seguida, apresentou-se a principal tecnologia *open source*: o DSpace-CRIS. Na Tabela 3.1 é efetuada uma comparação entre estas tecnologias. Finalmente, apresentaram-se 3 soluções tecnológicas (SQL Server, Spring, Angular) como alternativa à utilização de uma tecnologia já desenvolvida. A escolha destas soluções resultou de uma combinação entre conveniência de utilização de uma tecnologia já em uso por outros sistemas internos, experiência com as linguagens de programação envolvidas e o facto de algumas delas serem *open source*.

Tabela 3.1: Tabela de Comparação entre os CRIS

<b>Tecnologia</b>	<b><i>Open source</i></b>	<b>Compatibilidade com CERIF</b>	<b>Interoperabilidade com sistemas externos</b>
Pure	Não	Não	Sim
Converis	Não	Sim	Sim
DSpace-CRIS	Sim	Sim	Sim

## Capítulo 4

# Especificação

Neste capítulo especifica-se a solução proposta para o problema, incluindo a arquitetura geral e os modelos de dados. Embora existam dependências com outros subsistemas que, no conjunto, formam o sistema de informação interno, o trabalho desenvolvido incidiu, essencialmente, sobre 3 principais módulos: projetos, publicações e indicadores.

### 4.1 Arquitetura geral

Para a escolha das tecnologias a utilizar foram tidos em conta diversos fatores, tais como os sistemas de informação existentes ou o nível de conhecimento/experiência nas linguagens de programação envolvidas. Para a base de dados utilizou-se SQL Server, uma vez que é a tecnologia a ser já utilizada por outros sistemas no INESC TEC. Numa camada intermédia, que comunica diretamente com a base de dados, optou-se por utilizar Java (Spring com Hibernate). Esta camada intermédia disponibiliza *Web Services* à camada superior para operações CRUD sobre os dados. Esta camada superior utiliza a *framework* Angular para interface Web com o utilizador. A Figura 4.1 apresenta o conjunto das tecnologias selecionadas para cada camada de componentes, fornecendo uma visão geral da pilha tecnológica.

## Especificação

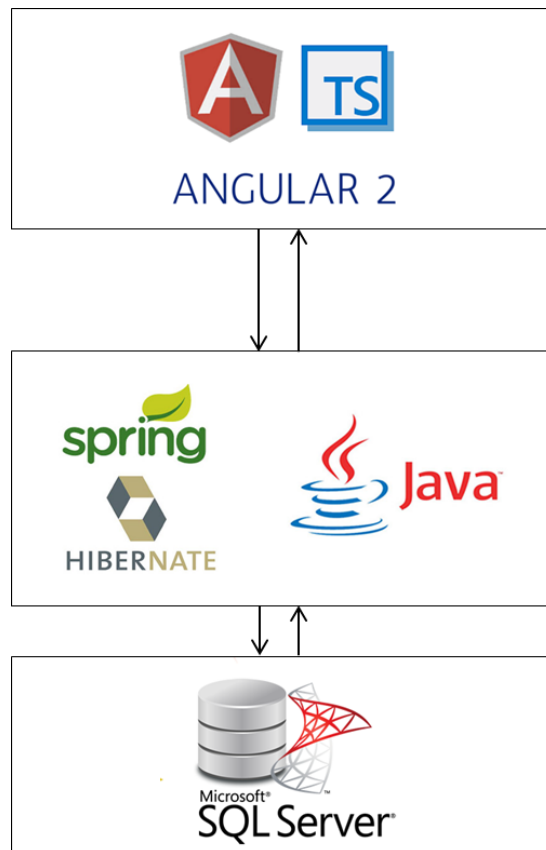


Figura 4.1: Pilha tecnológica

## 4.2 Projetos

Num CRIS, o conceito de projeto tem um papel central, já que praticamente toda a atividade de uma Unidade de Investigação se relaciona direta ou indiretamente com projetos de investigação, tendo sido, consequentemente, o principal foco de desenvolvimento.

### 4.2.1 Especificação

A informação básica dos projetos inclui identificadores internos, um nome curto, títulos e descrições em Português e Inglês, datas de início, fim previsto, fim efetivo e encerramento (se aplicável), um intervalo de valores de *Technology Readiness Level* (TRL)<sup>1</sup> nos quais o projeto se insere e um *Uniform Resource Locator* (URL) para um repositório de documentos. Um projeto tem também uma tipologia associada, que inclui uma fonte de financiamento e um intervalo de datas de validade. Para além disso, os projetos têm um tipo de destaque que pode estar ao nível

<sup>1</sup>HORIZON 2020 – WORK PROGRAMME 2014-2015 | General Annexes | G. Technology readiness levels (TRL), disponível em [https://ec.europa.eu/research/participants/data/ref/h2020/wp/2014\\_2015/annexes/h2020-wp1415-annex-g-trl\\_en.pdf](https://ec.europa.eu/research/participants/data/ref/h2020/wp/2014_2015/annexes/h2020-wp1415-annex-g-trl_en.pdf)

do projeto, do centro envolvido ou dum investigador envolvido. Os tipos de destaque incluem os valores "Confidencial", "Destaque", "Normal" e "Sucesso". O valor atribuído definirá o grau de visibilidade em diferentes áreas do sistema interno.

Os projetos estão organizados em *Work Packages* que possuem um identificador, um nome, uma descrição e datas de início e fim relativas ao início do projeto. Dentro de cada *Work Package* podem existir *deliverables*. Um *deliverable* é caracterizado por um título, uma descrição, um tipo e um *deadline* (que pode ser uma data fixa ou relativa ao início do projeto). Ao submeter um *deliverable*, é necessário registar a data de submissão, o autor e o URL para o documento.

Num projeto colaboram elementos de diferentes estruturas organizacionais que formam equipas com um coordenador. Cada colaborador tem uma função no projeto, uma dedicação (que representa a percentagem do seu tempo dedicada ao projeto), uma contribuição (que indica a percentagem do projeto para a qual contribuiu) e um intervalo de datas de participação.

Uma outra componente relevante é o planeamento de bolseiros. Para planear um projeto é necessário registar o número de bolseiros previstos por tipo de bolsa e intervalo de tempo. Com o decorrer do tempo, as vagas podem ser canceladas ou preenchidas como resultado de uma contratação. Neste último caso deve ficar registado qual o bolseiro associado e o intervalo de datas efetivo.

### 4.2.2 Modelo de dados

A Figura 4.2 apresenta o modelo de dados relativo ao módulo de projetos.

## Especificação

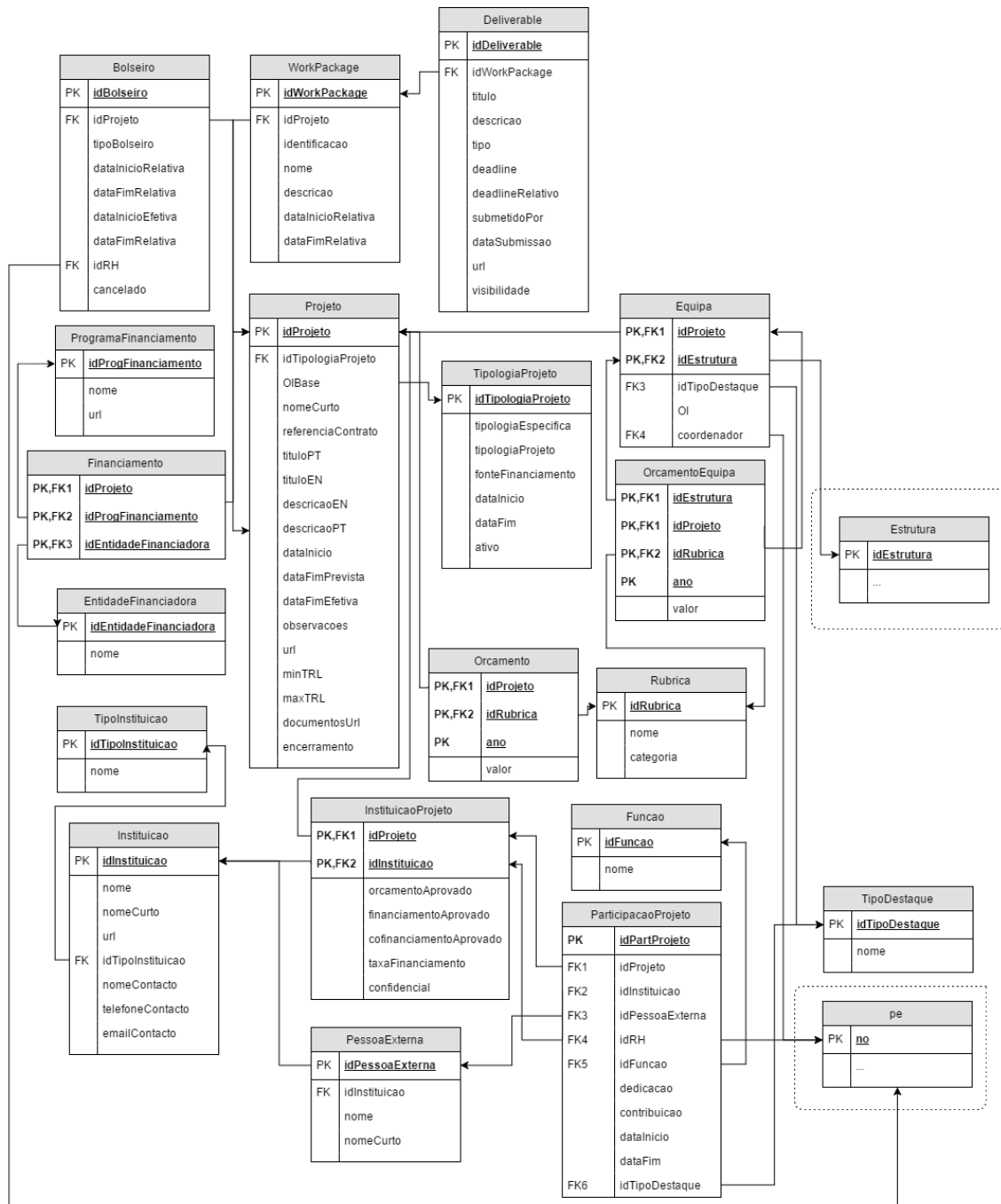


Figura 4.2: Modelo de dados de projetos

## 4.3 Publicações

Da participação em projetos são produzidos vários tipos de resultados, sendo as publicações um dos principais.

### 4.3.1 Especificação

O módulo de publicações deve incluir informação relativa ao próprio documento, como o título, *abstract*, data de publicação ou páginas. Para além disso, é necessário registar quais são os autores e a respetiva ordem ou a que projetos estão associados.

Um outro tipo de informação relaciona-se com a classificação das publicações, que engloba 3 vertentes: indexação em bases de dados de referência (WoS, Scopus ou DBLP); pontuação em quartis; classificação no *ranking* CORE<sup>2</sup>.

### 4.3.2 Modelo de dados

A Figura 4.3 apresenta o modelo de dados relativo ao módulo de publicações. A tabela central *Publicacao* regista a informação básica das publicações. A tabela *Publicacao\_Investigador* permite a ligação entre as publicações e os respetivos autores. A informação sobre a ordem do autor na publicação também é aqui registada.

Na tabela *ConferenciaCORE* são registadas os dados das conferências classificadas neste *ranking* juntamente com a classificação atribuída. Na tabela *PublicacaoCORE* é possível fazer a correspondência entre uma conferência classificada no CORE e uma publicação registada internamente. O método de correspondência - manual ou automático - também é guardado, assim como, no caso da correspondência manual, o utilizador que a realizou.

Na tabela *AreaCientifica* registam-se as diferentes áreas científicas que podem estar organizadas hierarquicamente. Na tabela *Quartil* guarda-se o quartil de uma determinada publicação numa área científica, de acordo com uma fonte (*Institute for Scientific Information* (ISI) ou Scopus).

---

<sup>2</sup>Computing Research & Education, disponível em <http://www.core.edu.au/>

## Especificação

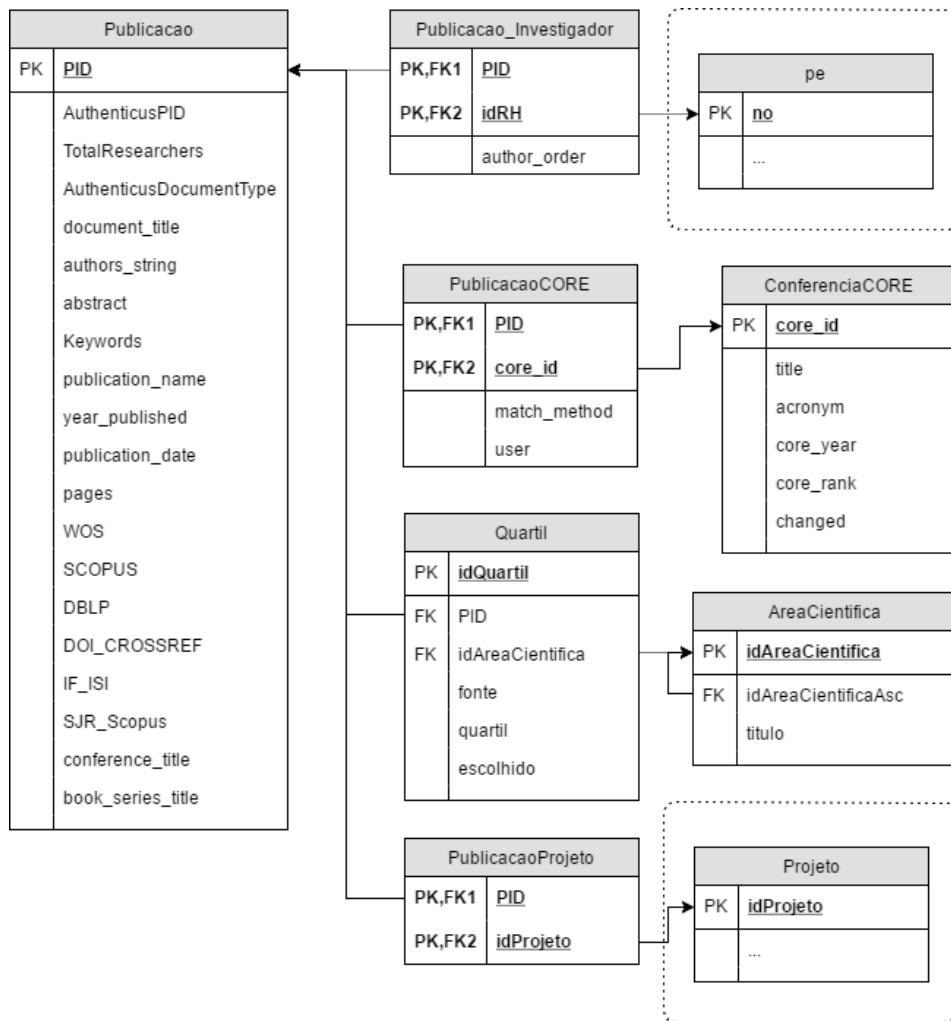


Figura 4.3: Modelo de dados de publicações

## 4.4 Indicadores

A atividade em projetos e publicações pode ser espelhada por indicadores que estão organizados por:

- Classificação do indicador (conjunto de campos que o caracterizam);
- Intervalo de tempo ao qual se referem (ano, semestre ou trimestre);
- Estrutura organizacional;
- Natureza do indicador (valor planejado, estimado, consolidado ou corrigido).



#### 4.4.1 Modelo de dados

A Figura 4.4 mostra o modelo de dados relativo ao módulo de indicadores implementado na base de dados em SQL Server. Numa primeira tabela - *Faceta* - são representados os tipos de indicadores e que incluem um tipo, um código, uma descrição, uma categoria e âmbito (que representam diferentes níveis de agregação), uma ordem para questões de apresentação, a unidade do indicador, dois campos de auditoria (data e autor da última alteração) e um intervalo de datas de validade. Na tabela *Indicador* são registados os valores propriamente ditos dos indicadores, registando ainda a informação do ano, período e natureza do indicador. Existem ainda 3 tabelas adicionais (*Natureza*, *Tipo*, *Periodo*) que definem o conjunto de valores possíveis dos atributos respetivos. Existe ainda a ligação ao módulo de estrutura, para indicar a que unidade organizacional um determinado indicador se refere.

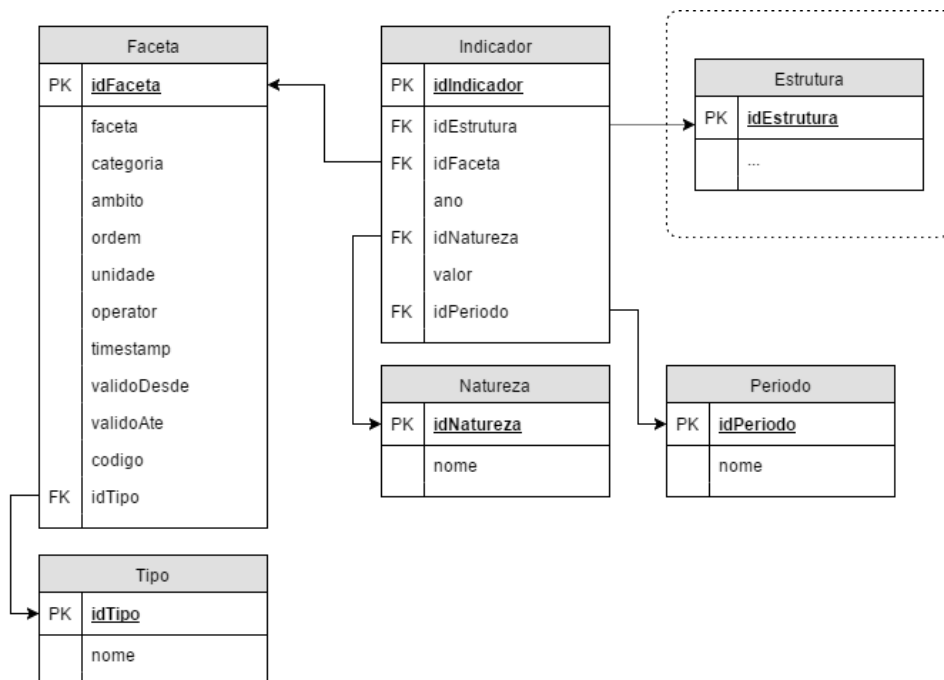


Figura 4.4: Modelo de dados de indicadores

## **4.5 Resumo**

Neste capítulo começou-se por apresentar uma visão geral da arquitetura proposta para a solução a nível de escolha de tecnologias. Em seguida, abordaram-se os conceitos relacionados com os 3 principais módulos - projetos, publicações e indicadores - de modo a compreender o tipo de informação que é necessário registar. Finalmente, apresentaram-se os modelos de dados sugeridos para responder a essas necessidades e que foram implementados na base de dados em SQL Server sobre a qual a interface Web em Angular opera.

## Capítulo 5

# Implementação

Este capítulo tem como objetivo apresentar com maior detalhe a plataforma desenvolvida, incluindo o servidor Spring, a interface Web em Angular e cenários típicos de utilização. Priorizou-se o desenvolvimento do módulo de projetos, avançando-se também parte do módulo de publicações. Finalmente, iniciou-se também o desenvolvimento do *Web Service* de exportação em CERIF XML.

### 5.1 Servidor Spring

No servidor de Spring foram desenvolvidos vários *Web Services* que fornecem uma interface entre o Angular e o acesso a dados:

- *getProjetos()* - Devolve o conjunto de projetos existentes com a informação básica de cada um;
- *getProjeto()* - Devolve toda a informação associada a um projeto;
- *getPessoas()* - Permite obter os dados básicos dos recursos humanos;
- *getCentros()* - Usado para obter a lista de centros atualmente válidos;
- *getTiposDestaque()* - Devolve o conjunto de valores possíveis para os tipos de visibilidade de um projeto;
- *getTiposAtividadeDominante()* - Devolve o conjunto de valores possíveis para a atividade dominante de um projeto;
- *editProjeto()* - Permite a edição dos dados de um projeto;
- *getCERIF()* - Gera um ficheiro no formato CERIF XML com a informação na base de dados;
- *getPublicacoes()* - Devolve o conjunto das publicações existentes.

## 5.2 Projetos

No módulo de projetos são apresentadas 3 páginas principais: uma primeira para pesquisa e listagem de projetos; uma segunda página para visualização dos dados associados a um projeto; e, finalmente, uma terceira página para edição num formato idêntico à de visualização. As permissões de acesso a cada uma destas páginas são determinadas através de um pedido ao servidor de LDAP do INESC TEC, mas que está fora do âmbito desta dissertação.

### 5.2.1 Página de pesquisa

Na página de pesquisa são listados os projetos que cumprem os critérios especificados (nome curto, título ou um identificador interno). Ao carregar num elemento dessa lista, o utilizador navega para a página de visualização do projeto correspondente. Na Figura 5.1 é possível ver a página de pesquisa desenvolvida.

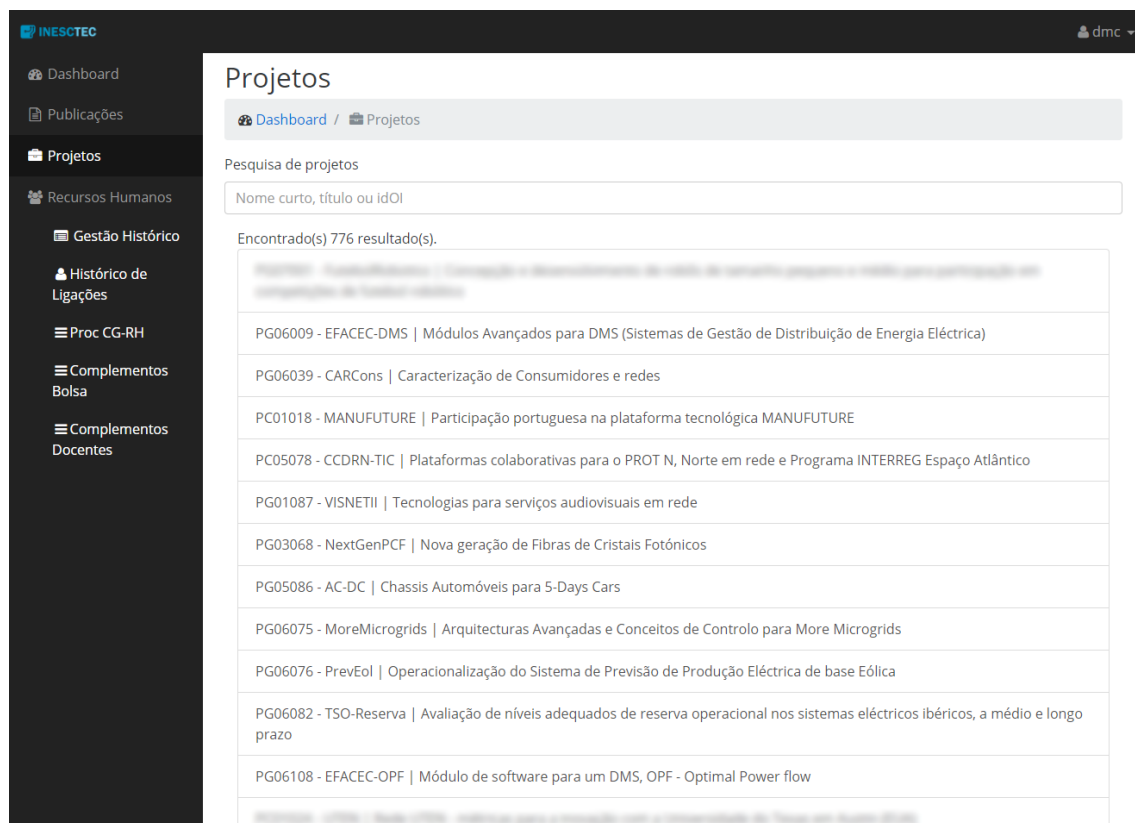
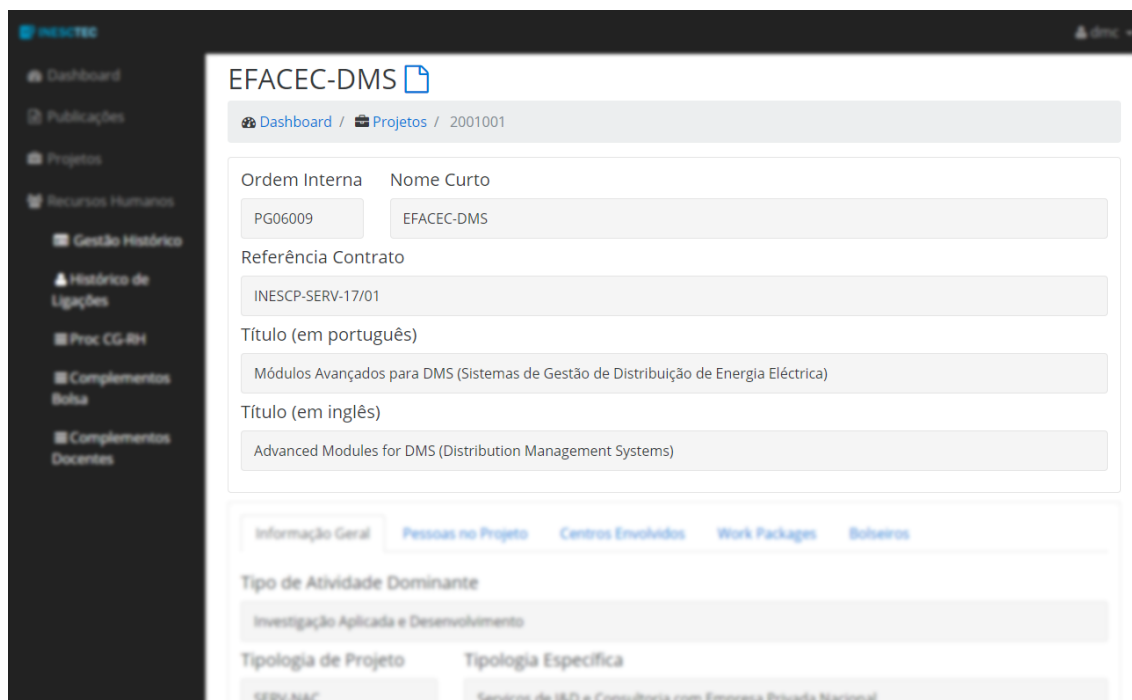


Figura 5.1: Página de pesquisa de projetos

### 5.2.2 Página de visualização

A página de visualização está dividida em 2 secções: uma, no topo, com dados de identificação do projeto como o nome curto ou títulos em Português e Inglês; e outra, por baixo, com outro tipo de dados organizados em separadores. Quando um campo previsto não está preenchido é omitido da apresentação. A primeira secção desta página pode ser vista na Figura 5.2.



The screenshot displays the 'EFACEC-DMS' project identification page. The sidebar on the left contains navigation links: Dashboard, Publicações, Projetos, Recursos Humanos, Gestão Histórico, Histórico de Ligações, Proc. CG-RH, Complementos Bolsa, and Complementos Docentes. The main content area is titled 'EFACEC-DMS' and includes a breadcrumb trail: Dashboard / Projetos / 2001001. The project details are organized into sections: 'Ordem Interna' (PG06009) and 'Nome Curto' (EFACEC-DMS); 'Referência Contrato' (INESCP-SERV-17/01); 'Título (em português)' (Módulos Avançados para DMS (Sistemas de Gestão de Distribuição de Energia Eléctrica)); and 'Título (em inglês)' (Advanced Modules for DMS (Distribution Management Systems)). Below these details is a tabbed interface with tabs for 'Informação Geral', 'Pessoas no Projeto', 'Centros Envolvidos', 'Work Packages', and 'Bolsistas'. The 'Informação Geral' tab is active, showing 'Tipo de Atividade Dominante' (Investigação Aplicada e Desenvolvimento), 'Tipologia de Projeto' (SERV NAC), and 'Tipologia Específica' (Serviços de I&D e Consultoria com Empresa Privada Nacional).

Figura 5.2: Identificação do projeto

## Implementação

No primeiro separador da secção inferior, é apresentada a informação geral do projeto que inclui o tipo de atividade dominante, tipologias, financiamento e datas de início, fim previsto e fim efetivo. Existe também um conjunto de estados do projeto determinados a partir destas datas. A Figura 5.3 apresenta a vista deste separador.

The screenshot shows the 'Informação Geral' (General Information) tab of a project form in the INESC TEC system. The form contains the following fields:

- Referência Contrato:** INESCP-SERV-17/01
- Título (em português):** Módulos Avançados para DMS (Sistemas de Gestão de Distribuição de Energia Eléctrica)
- Título (em inglês):** Advanced Modules for DMS (Distribution Management Systems)
- Tipologia de Projeto:** SERV-NAC
- Tipologia Específica:** Serviços de I&D e Consultoria com Empresa Privada Nacional
- Fonte de Financiamento:** SERV - NAC
- Estado Atual:** Ativo - a confirmar
- Estado Fim Período:** Ativo
- Estado Progresso:** Continuação
- Data de Início:** 2001-04-15

At the bottom right, there is an 'Editar' (Edit) button.

Figura 5.3: Informação geral do projeto

O segundo separador fornece uma vista dos investigadores que participam no projeto, juntamente com a função desempenhada, dedicação, contribuição e intervalo de datas de atividade de cada um. Existe ainda a informação do estado de validação de cada linha por parte do responsável. É possível observar na Figura 5.4 um exemplo com dados de teste deste separador.

The screenshot shows the 'Pessoas no Projeto' (People in the Project) tab. It contains a table with the following data:

Nome	Função	Dedicação (%)	Contribuição (%)	Data de Início	Data de Fim	
Elvio Accinelli		100		2016-06-01	2019-05-31	<input type="checkbox"/>

At the bottom right, there is an 'Editar' (Edit) button.

Figura 5.4: Investigadores no projeto

## Implementação

Num terceiro separador, registam-se os centros envolvidos no projeto. No quarto separador são mostrados os *Work Packages* do projeto com a respetiva identificação, o nome, duração, descrição e conjunto de *deliverables*. Cada *deliverable* tem um título, um tipo (que é um campo livre com sugestão de valores pré-definidos), uma descrição e uma data de entrega. Também é apresentado o estado do *deliverable* e, no caso de ter sido submetido, o autor, a data de submissão e o URL do documento (se aplicável).

No último separador, é apresentada a lista de bolsеiros no projeto. Cada linha inclui o tipo de bolsa, e as datas de início e fim - previstas, no caso de ser um vaga não preenchida; efetivas, no caso de existir um colaborador associado, registando-se também neste caso o investigador.

### 5.2.3 Página de edição

A página de edição de um projeto tem o mesmo formato que a de visualização. A título de exemplo, apresentam-se nas Figuras 5.5, 5.6 e 5.7 a interface desenvolvida para a edição de, respetivamente, centros, *Work Packages* e bolsеiros.

Centro	Coordenador	Ordem Interna	Divulgação
LIAAD	Alberto Pinto	PL01021	Normal

Figura 5.5: Edição de centros envolvidos no projeto

Identificação: Nome

Duração: Mês de Início 1 Mês de Fim 12

Descrição:

Deliverables:

Título	Tipo	Descrição	Deadline	Estado
--------	------	-----------	----------	--------

Figura 5.6: Edição de *Work Packages* do projeto

## Implementação

**Dynamics2**

Dashboard / Projetos /

Ordem Interna: PL01021

Referência Contrato: PTDC/MATNAN/6890/2014

Título (em português): Dinâmica, optimização e mod

Título (em inglês): Dynamics, optimization and m

**Editar/Adicionar Bolsheiro**

Tipo de Bolsa: BGCT - Gestão de Ciência e Tecnologia

Intervalo de datas: ☐ Relativo ao início do projeto ☒ Associar colaborador

Data de Início:  Data de Início

Data de Fim:  Data de Fim

Colaborador:  Colaborador

Informação Geral | Execução | Pessoas no Projeto | Centros Envolvidos | Work Packages | **Bolsheiros**

Tipo de Bolsa	Início	Fim	Colaborador Associado
---------------	--------	-----	-----------------------

Figura 5.7: Edição de bolsheiros do projeto

## 5.3 Publicações

No módulo de publicações, é possível listar publicações de acordo com alguns critérios: ano, tipo de documento e indexação. Os dados apresentados incluem o título, os autores, o tipo de publicação, o ano e alguns atributos de classificação. Os atributos de classificação podem ser: indexado (ISI, Scopus ou *Digital Bibliography & Library Project* (DBLP)), quartil (1º a 4º) e *ranking* no CORE (no caso das conferências). A Figura 5.8 apresenta a interface desenvolvida para este módulo.

**INESCTEC**

Dashboard / Publicações

Ano: 2016 Tipo de documento: Abstract Apenas indexados ☒

Total: 4

AuthPID	Título	Autores	Tipo de Publicação	Ano	Classificação
P-00M-571	Exercise Improves Depression: Gold Strategies to Treatment Adherence?	Vasconcelos raposo, J	Abstract	2016	1
P-00M-8X2	Locoregional management of hereditary breast cancer: opportunities against timing	Cardoso, MJ	Abstract	2016	1
P-00M-8X3	When and how to reconstruct considering long-term effects	Cardoso, MJ	Abstract	2016	1
P-00M-FXF	Predicting Clinical Outcomes of Gastric Endoscopic Submucosal Dissection Using a Bayesian Approach	Rodrigues, PP	Abstract	2016	1

Figura 5.8: Módulo de publicações



## 5.4 Exportação em CERIF XML

No servidor de Spring foi também desenvolvido um *Web Service* que permite a exportação dos dados existentes no sistema de informação. Para isso, foi necessário mapear a informação aí existente para os elementos correspondentes no modelo de dados CERIF. Começou-se por mapear os elementos que tinham uma correspondência direta entre os dois modelos. Contudo, esses elementos constituem apenas um pequeno subconjunto dos dados que se pretendem exportar. Por um lado, o formato CERIF XML tem uma sintaxe própria que implica algum tratamento da informação existente. Por outro, este formato não prevê todas as necessidades e todos os detalhes do sistema científico nacional. Então, procedeu-se em seguida ao mapeamento de outros elementos que não tinham correspondência direta.

O XML gerado cobre atualmente os dados básicos dos projetos: identificador interno, datas de início e fim, acrónimo, títulos (em Português e Inglês) e resumos (também em Português e Inglês). Relativamente às publicações, os dados devolvidos são: o identificador, o número de páginas, *International Standard Serial Number* (ISSN), título e resumo (em Inglês).

## 5.5 Resumo

Neste capítulo apresentou-se a plataforma desenvolvida com maior detalhe. Começou-se por descrever os *Web Services* desenvolvidos na camada intermédia com recurso à tecnologia Spring. Em seguida, apresentou-se a interface Web dos módulos de projetos e publicações, bem como as suas principais funcionalidades. Finalmente, foi referida a funcionalidade de exportação em CERIF XML, incluindo uma descrição do processo envolvido na geração do documento neste formato.

## Implementação

## Capítulo 6

# Conclusões e Trabalho Futuro

Neste capítulo são apresentadas as principais conclusões relativas aos objetivos atingidos e ao projeto em geral, bem como o trabalho futuro a ser desenvolvido.

### 6.1 Conclusão e Satisfação dos Objetivos

A tarefa de desenvolver um sistema de informação de investigação, tendo em conta as questões de interoperabilidade com outros sistemas, não é trivial. A principal dificuldade observada prende-se com o conflito entre a necessidade de cumprir com os requisitos necessários à Unidade de Investigação, com todos os detalhes e especificidades que englobam e a conveniência da adoção de um formato comum para representação da informação.

O trabalho desenvolvido incluiu a especificação da arquitetura dos subsistemas que constituem o CRIS do INESC TEC, focando-se em 3 módulos principais: projetos, publicações e indicadores. Relativamente à implementação, priorizou-se o módulo de projetos, avançando-se também nos módulos de publicações e exportação no formato CERIF XML. Deste modo, pode-se concluir que os principais objetivos (ver Capítulo 1) foram atingidos, existindo, contudo, diversas sugestões de trabalho futuro.

### 6.2 Trabalho Futuro

O trabalho desenvolvido até à data é um subconjunto da solução proposta a longo prazo. Assim, enumeram-se em seguida os principais passos de desenvolvimento futuro previstos.

- O XML gerado pelo *Web Service* de exportação em CERIF representa apenas uma parte da informação existente, já que teve uma prioridade inferior. Contudo, pretende-se expandir o conjunto de dados cobertos pelo XML de modo a melhor representar o conteúdo da base de dados;

## Conclusões e Trabalho Futuro

- Pretende-se desenvolver uma interface para o módulo de indicadores, com operações CRUD sobre os dados;
- Finalmente, pretende-se permitir o carregamento de novas publicações.

# Referências

- [HJD<sup>+</sup>14] Nikos Houssos, Brigitte Jörg, Jan Dvořák, Pedro Príncipe, Eloy Rodrigues, Paolo Manghi e Mikael K. Elbæk. OpenAIRE Guidelines for CRIS Managers: Supporting Interoperability of Open Research Information through Established Standards. *Procedia Computer Science*, pages 33 – 38, 2014.
- [Jö10] Brigitte Jörg. CERIF: The Common European Research Information Format Model. *Data Science Journal*, 9:CRIS24–CRIS31, 2010. doi:10.2481/dsj.CRIS4.
- [Jef10] Keith Jeffery. The CERIF Model As the Core of a Research Organization. *Data Science Journal*, 9:CRIS7–CRIS13, 2010. doi:10.2481/dsj.CRIS2.
- [PBMM14] David T. Palmer, Andrea Bollini, Susanna Mornati e Michele Mennelli. DSpace-CRIS@HKU: Achieving Visibility with a CERIF Compliant Open Source System. *Procedia Computer Science*, 33:118 – 123, 2014. doi:http://dx.doi.org/10.1016/j.procs.2014.06.019.
- [vLvWW16] Thed N. van Leeuwen, Erik van Wijk e Paul F. Wouters. Bibliometric analysis of output and impact based on CRIS data: a case study on the registered output of a Dutch university. *Scientometrics*, 106(1):1–16, 2016. URL: <http://dx.doi.org/10.1007/s11192-015-1788-y>, doi:10.1007/s11192-015-1788-y.